How the statistical model works

This document sets out how our statistical model works.

Model assumptions

The aim of this analysis is to describe timecourse data about the density of CHO cell cultures that were given treatments that induce cell death. Some cultures were given genetic interventions that aim to make them apoptosis-resistant, either by reducing the rate at which they die or by extending the period before they start to die. We would like to know which interventions have the most effect, and in what way.

We assume that in this scenario the cells exist in four states:

- R replicative, growing at a rate of $\mu R(t)$, where t is the current time and $R(t)$ is the current density of replicative cells
- Q_a growth arrest, transferring from normal at a rate of $kqR(t)$
- Q_c death committed, transferring from growth arrest at a rate of $kqR(t-\tau)$, where $tau > 0$ represents the delay between growth arrest and death commitment.
- dead, transferring from death committed at a rate of $kdQ_c(t)$, where $Q_c(t)$ is the density of death-committed cells at time t.

These assumptions define a system of ordinary differential equations (specifically "delay differential equations") that can be solved analytically, so that the density at a given time can be found as a function of the parameters μ, τ kq, kd and the initial density of replicative cells $R0$.

We have measurements of the total cell volume, i.e. $R(t) + Q_a(t) + Q_c(t)$ at several time points, for cell cultures with the following structure:

- 9 genetic designs, comprising 7 genetic interventions and two control designs.
- Between 1 and 4 clones implementing each design
- Two technical replicates for each clone.

Replicates of the same clone are biologically identical, though we expect some variation in measurements due to the experimental conditions. Clones with the same design are expected to be similar, with some degree of clonal variation that may differ depending on the design. There is no prior information distinguishing the designs from each other, or distinguishing clones with the same design.

Given these assumptions a multi-level Bayesian statistical model is appropriate.

We used the following measurement model:

$$y \sim lognormal(\log(\hat{y}(t, R0, \mu_r, \tau_r, kq_r, kd_r)), \sigma)$$

where

- $R0$, μ_r, τ_r, kd_r and kd_r are vectors of replicate-level parameters
- σ is an unknown log-scale error standard deviation
- t is a vector of known measurement times (one per measurement)
- $\hat{y}$ is a function mapping parameter configurations to densities, under the delay differential equation assumptions laid out above

The parameters μ_r, τ_r, kd_r and kd_r vectors are treated as compounds of design-level means and a clone level residuals, i.e.

$$\begin{aligned}
\mu_r &= \mu + a_\mu \\
\tau_r &= \tau + a_\tau \\
kq_r &= kq + a_{kq} \\
kd_r &= kd + a_{kd}
\end{aligned}$$

where μ, τ, kd and kq are design-level parameters and e.g. a_τ is a vector of clone-level residuals.

In order to accommodate uncertainty as to the level of clonal variation, we use hierarchical prior distributions for the clone level parameters, e.g.

$a_\mu \; normal(0, sd_{a_\mu})$

Other unkowns have informative prior distributions based on scientific knowledge.